Commentary

The Potential and Limitations of Data From Population-Based State Cancer Registries

Jorge N. Izquierdo, DDS, PhD, MPH, and Victor J. Schoenbach, PhD

A B S T R A C T

Cancer incidence varies markedly among states because of population heterogeneity regarding risk, genetic, and demographic factors. Population-based cancer registries are essential to monitoring cancer trends and control.

The Centers for Disease Control and Prevention and the North American Association of Central Cancer Registries, through the National Program of Cancer Registries, are helping state registries generate more and better data nationwide. The National Program of Cancer Registries has supported the enhancement of 36 registries and the creation of 13 new registries in 45 states, 3 territories, and the District of Columbia, providing national standards for completeness, timeliness, and quality; financial support; and technical assistance.

Users must be aware of diverse issues that influence collection and interpretation of cancer registry data, such as multiple cancer diagnoses, duplicate reports, reporting delays, misclassification of race/ethnicity, and pitfalls in estimations of cancer incidence rates. Attention to these issues and intense use of the available data for cancer surveillance will enable maximum societal benefit from the emerging network of population-based state cancer registries. (*Am J Public Health*. 2000;90: 695–698)

Population-based state cancer registries collect, classify, consolidate, and link information on new cancer cases from hospital reports, medical records, pathology reports, hospital discharge abstracts, and death certificates. Through a time- and resource-intensive process,[1] the registries generate data of vast surveillance potential for identifying patterns and trends in various population groups, in different geographic areas, and over time; orienting prevention efforts and health care planning; supporting epidemiologic, biomedical, and health services research; and framing public health policy.[2–8] Patterns of cancer incidence and mortality vary markedly between and within states and regions in the United States[9–11] and across countries,[12] reflecting population heterogeneity in regard to demographic and genetic composition as well as exposure to environmental and behavioral risk factors.[13–16] Thus, a national system of population-based cancer registries is essential to monitor state- and local-level cancer patterns and trends and to orient cancer prevention and control activities.[8]

Despite the importance of local cancer data in developing and evaluating control measures, 10 states had no cancer registry in 1990, and many existing ones lacked financial or technical resources to collect complete, accurate, and timely data of requisite quality. In 1992, Congress passed the Cancer Registries Amendment Act (Pub L 102-515), establishing the National Program of Cancer Registries. This legislation authorized the Centers for Disease Control and Prevention (CDC) to provide funds to states and territories to improve existing cancer registries; plan and implement new registries; develop model legislation and regulations for states to enhance the viability of registry operations; set standards for data completeness, timeliness, and quality; provide training for registry personnel; and help establish a computerized reporting system.[17] To ensure complete and timely reporting in each state, the federal statute requires state-specific laws authorizing the cancer registry along with 8 categories of regulations intended to (1) require mandatory reporting of newly diagnosed cancer cases by hospitals and other health care facilities, (2) require reporting of cancer cases by physicians and other practitioners, (3) guarantee access by the state cancer registry to all medical records of persons with cancer, (4) require use of standardized reporting formats, (5) ensure confidentiality of cancer case data, (6) allow use of confidential data by researchers, (7) authorize studies using cancer registry data, and (8) protect persons complying with the law from liability.[18]

When the CDC created the National Program of Cancer Registries, 37 states had laws authorizing state cancer registries, 14 had all enabling regulations in place, and only 9—California, Indiana, Minnesota, New Jersey, New York, Ohio, Rhode Island, Vermont, and Washington—had both components.[18] During 1997, the program expanded by adding 6 registries (Connecticut, Delaware, Tennessee, Puerto Rico, Virgin Islands, and Republic of Palau), increasing the number of supported registries to 49 (45 in states, 3 in territories, and 1 in the District of Columbia).[17] By 1999, 45 states and the District of Columbia had authorizing legislation, and 39 states had enabling regulations.

The CDC set national standards for completeness (95% unduplicated malignant cases), timeliness (reports within 6 months of diagnosis), and quality[19] and, in fiscal 1997, allocated $22.3 million ($24 million in 1998 and 1999[17]) to enable the National Program

The authors are with the School of Public Health, University of North Carolina at Chapel Hill.

Requests for reprints should be sent to Jorge N. Izquierdo, DDS, PhD, MPH, School of Public Health, University of North Carolina, Chapel Hill, NC 27599-7400 (e-mail: jorge_izquierdo@unc.edu).

This commentary was accepted January 3, 2000.

of Cancer Registries to help states enhance registries and advance the agenda for a centralized database of cancer incidence in the United States. A key player at all levels of this development process is the North American Association of Central Cancer Registries, which, through cooperative agreement and federal contract, provides guidance, training, technical assistance, and quality audits.[20]

Pitfalls in State Cancer Registry Data

While the expanding population-based state cancer data hold great promise for cancer surveillance, various considerations influence their validity and usefulness.

Cases vs Persons

Cancer registries record cancer cases, not patients. Because a patient may have multiple primary cancers, the same person can appear more than once in a registry database. A person with primary cancers in the lung and breast (or both breasts) is considered as 2 cases; a person with primary cancers in the tongue, cheek, and palate is considered as 3 oral cancer cases. Because numbers of patients, rather than numbers of cases, are often necessary for health care planning and research design, such figures should be routinely provided, or simple procedures to reliably estimate them from numbers of cases need to be developed.

Duplicate Reports

When the same cancer is diagnosed and reported by more than 1 provider (as when a patient obtains a second opinion), the registry must consolidate such multiple reports into a single case record. With information coming from diverse sources and at different times, the "unduplication" process is often tedious and uncertain, because a second report could represent a recurrence of the original tumor, a new primary tumor in the same organ, or a duplicate report of the same case (with the same or a different diagnosis).

Reporting Delays

Notwithstanding regulations in regard to timely reporting of new cases, late reports are received, for diverse reasons. Providers may not comply with deadlines; institutions may discover additional cases after sending in their lists of cases; and unreported cases may be detected from death certificates. Forwarding of reports for residents receiving cancer care in another state introduces an additional step. Once received by the registry, each report requires review for potential duplicates. For these reasons, registry administrators can find themselves having to repeatedly postpone the release of data or make changes in a data set after it has been released. Both of these situations are sources of frustration for investigators as well as registry personnel, who are committed to accuracy, precision, consistency, and timeliness.

Race/Ethnicity

Although the inclusion of race/ethnicity in official databases raises numerous issues, such information is necessary for research, monitoring, and policy-making regarding cancer in populations whose life situations, culture, and health care experience have differed greatly from those of the majority population throughout US history.[21–26] Comparisons across groups can suggest avenues for research in cancer etiology, prevention, and outcomes and are required for assessing disparities in care. However, difficulties arise in coding race/ethnicity, because concepts, perceptions, methodologies, and the populations themselves change over time.

Because a uniform method of assessing and coding race/ethnicity does not exist,[19] sources of and coding methods for race/ethnicity data represent important information for data users. The method adopted in the US census is self-identification: the individual designates his or her own race/ethnicity. By contrast, race/ethnicity information in medical records usually represents the perception of physicians, nurses, or clerks. Both patients' and providers' criteria and motivations for selecting one designation or another vary across situations.

In the case of some groups, coding is particularly susceptible to inaccuracy. For example, American Indians are frequently classified as White by health care workers. A recent report from the California Cancer Registry showed that when its 1988 to 1992 database was linked with the Indian Health Service database, 1478 American Indian cancer cases were identified, 844 (57.1%) of which had been previously misclassified as non–American Indian cases.[27] During its first 3 years of statewide operation, 41% (1990), 26% (1991), and 22% (1992) of reports to the North Carolina Central Cancer Registry lacked information on "Hispanic origin," which is problematic, especially in analyses of trends.

Cancer Rates

Cancer registry data are primarily used to estimate cancer incidence rates. These estimates are vulnerable to pitfalls regarding numerators, denominators, and methods of estimation. Changes in numbers of cases, misclassification, and missing data can substantially distort numerators and, consequently, rates. For example, correction of the earlier-cited misclassification of American Indians in California resulted in an average annual age-adjusted incidence rate (211.0/100 000 for all sites combined) more than twice the uncorrected figure (89.6/100 000).[27]

Likewise, although population counts, estimates, and projections from the US Bureau of the Census and state agencies generally provide adequate denominators for estimating rates, undercounts and inaccurate projections affect rate estimates, particularly for minority groups.[28] The problem of undercounting inner-city African Americans has been recognized for years. A different situation with a similar effect is the inaccuracy in projections for rapidly increasing populations, such as Hispanics/Latinos in North Carolina, Virginia, and other southeastern states during the 1990s,[29] accentuated by the exclusion of migrant workers and their families, predominantly Hispanics/Latinos.[30] Even when both numerators and denominators are accurate, the comparability of age-standardized cancer rates may be compromised when, owing to the small sizes of minority populations in many states, standardization must be carried out via the "indirect" instead of the "direct" method.[31,32]

The 1994 Gap

The National Program of Cancer Registries designated 1995 as the year for participating registries to adopt CDC standards for data collection and processing. Hence, beginning in 1995, treatment-related data had to be collected by registries that had never done so, had to be submitted by a uniform deadline, and had to meet completeness and quality requirements. In many state registries, achieving this goal required nearly all available resources. The strategy promoted a great leap forward toward uniform state cancer incidence databases but had an unfortunate side effect. To meet national standards for the 1995 data, many registries had to defer processing their 1994 data (in some cases, these data still had not been officially released as of this writing). The resulting gap interferes with analyses involving multiple years.

Facilitating Research vs Protecting Privacy

With the growing use of geographical information systems, the conflict between confidentiality and data accessibility has become increasingly problematic. For example, be-

cause individual-level data on socioeconomic status (SES) are generally not reported to registries, investigators have used small-area data from the US census (median family income, median educational level) as indicators of SES. More types of data (e.g., location of health care facilities, pollution levels, soil characteristics) can be analyzed through geographical information systems, and thus the value of locational data for cases has grown considerably. Yet, releasing locational data will often enable identification of individual patients, thereby compromising confidentiality.

Although the regulations required under the Cancer Registries Amendment Act guarantee confidentiality of patient information and access to data for research purposes, the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program and many state registries have been reluctant to release detailed locational information. Similarly, when researchers attempt to contact a patient for interviews or to obtain consent to examine medical records, their request may be refused by the patient's physician or may prompt hospital registrars to complain to the state registry.

Rapid Case Reporting

The integration of population-based epidemiology and molecular biology provides new opportunities for cancer etiological studies but also increases the need for rapid identification of cases to enable the collection of blood and tissue specimens promptly after diagnosis. However, case reports typically arrive at the registry several months after diagnosis. To circumvent this delay, investigators from the Carolina Breast Cancer Study and the North Carolina Central Cancer Registry implemented a rapid case reporting system for hospitals in the study area that enabled the identification of newly diagnosed breast cancer cases in less than 1 month.[33,34] Nevertheless, rapid case reporting implies additional work that some registries are not prepared to deal with, even when funds are allocated for extra personnel.

Carcinoma in situ

Carcinoma in situ poses something of a dilemma for cancer registries. Because, by definition, carcinoma in situ is not invasive, it is not classified as a "real" cancer and is therefore not included in computations of incidence rates (except for bladder cancer, which has a distinctive transitional epithelium). Although they often collect data on carcinoma in situ (except for carcinoma in situ of the cervix), most cancer registries do not routinely analyze or publish these data. Because of the high risk of malignant transformation, systematic surveillance of carcinoma in situ could facilitate studies of the etiopathogenesis of progression to invasive cancer and research on prognostic factors.[35,36]

Also, because in situ lesions represent an early step in the natural history of cancer, case–control studies of carcinoma in situ have advantages in regard to investigating risk factors that are less detectable with the passage of time owing to recall problems and difficulties in obtaining medical records from the distant past.[35] Because the number of cases of carcinoma in situ is substantial (e.g., 67 255 of the cases [nearly 12%] reported by the California Cancer Registry from 1988 through 1992[37]), systematic collection, analysis, and reporting of such data can contribute to cancer control research.[35,36]

Perspectives

The expansion/enhancement of state cancer registration supported by the National Program of Cancer Registries is an important step in the development of an infrastructure for cancer surveillance, prevention/control, and research. Currently, the main challenge is to continue nurturing the ongoing process: further expanding to cover all states, providing training and audits to reach required standards in each registry, finding solutions to issues such as those described here, linking cancer registry databases with vital statistics (mortality) and administrative (e.g., Medicare) databases, and periodically evaluating registry quality through a systematic certification process (presently conducted, on a voluntary basis, by the North American Association of Central Cancer Registries, with National Program of Cancer Registries funding). Through these efforts, the heterogeneously developed registries will gradually evolve into an efficient, mutually compatible, nationwide network of population-based state cancer registries.

This emerging network should efficiently complement the existing registries, especially the SEER program. SEER has collected cancer data for more than 25 years; it currently covers 14% of the US population and has recently been linked to Medicare databases.[38] Within the next decade, the state cancer registry network and SEER together will collect cancer incidence data on approximately 97% of the US population,[17] laying the foundation for a national system of cancer surveillance.

Availability of more and better data will lead to greater use of these data, especially among cancer epidemiology and clinical oncology research and training programs and public health practitioners. Reciprocally, this data use, nurtured by new scientific knowledge and technological advances, will generate both demands for additional high-quality data and the potential for conflict. For example, the viability of investigations combining epidemiologic and molecular data depends on the ability of investigators to avoid overburdening providers and registries, as well as on the attitudes toward research of each cancer registry, hospital, laboratory, and health professional involved. Development of mutual understanding and trust among investigators, clinicians, and registry personnel is vital.

A key challenge is to balance the needs of investigators, patients, health care providers, and registries to simultaneously ensure confidentiality and the scientific usefulness of data. Research institutions and cancer registries need to develop regulations better defining the conditions under which qualified investigators can gain access to data with the potential to reveal identities. Creative solutions to the conflict between protecting patients' privacy and enabling high-quality health research will help. In achieving consensus on the need for public-use data files of national cancer data, participants in a recent North American Association of Central Cancer Registries workshop concluded that no one file can meet the needs of all users, maintain confidentiality, and be sufficiently flexible to accommodate the wide range of potential uses. They recommended the development of 4 separate files covering a range of user needs and balancing issues of confidentiality, flexibility, and content.[39]

An efficient national cancer registry system is only an intermediate objective; the ultimate goals are to further understand and control cancer. Advances in cancer registration should be coupled with a nationwide effort to foster the systematic investigation of cancer patterns and trends by states, regions, and subpopulations so that, over the next decade, such information will nurture the design, implementation, and evaluation of interventions in each state. Although a national surveillance system is not mentioned in the 1992 Cancer Registries Amendment Act, the National Program of Cancer Registries could proactively collaborate with diverse cancer-related organizations to jointly analyze data, help define priorities, suggest prevention/control interventions, update the research agenda, and inform policymakers and public opinion.

Recently, a report by the National Coordinating Council for Cancer Surveillance, created in 1995 to facilitate collaboration among organizations interested in cancer surveillance, introduced the notion of a "nationally coordinated system that achieves comparability among current programs while

remaining relevant, maintaining high quality, and avoiding extraordinary costs."[40(p1283)] It is through its involvement in such multi-institutional collaborative endeavors that the emerging network of state cancer registries can play a critical role as a foundation for cancer surveillance, prevention/control, and research in the United States. □

Contributors

J. N. Izquierdo and V. J. Schoenbach developed the plan for the manuscript, which was drafted by J. N. Izquierdo. Both authors participated in multiple revisions, including that resulting in the final manuscript.

Acknowledgments

This work was supported by the University of North Carolina Minority Cancer Control Research Program under Public Health Service awards P30 CA16086-14S1 and R01 CA64060 to the Lineberger Comprehensive Cancer Center.

We acknowledge the helpful assistance of the North Carolina Central Cancer Registry, especially its former director, Dr Tim E. Aldrich, who kindly supported this work and provided feedback during the preparation of the manuscript. We also acknowledge the helpful assistance of the Cecil G. Sheps Center for Health Services Research.

References

1. Hutchison C, Roffers S, Fritz A. *Cancer Registry Management: Principles and Practice*. Lenexa, Kan: Kendall/Hunt Publishing Co; 1997.
2. Storm HH. Cancer registries in epidemiologic research. *Cancer Causes Control*. 1996;7:299–301.
3. Stroup N, Zack M, Wharton M. Sources of routinely collected data for surveillance. In: Teutsch S, Churchill R, eds. *Principles and Practice of Public Health Surveillance*. New York, NY: Oxford University Press Inc; 1994:31–85.
4. Aldrich T. Research uses of central registries. In: Menck H, Smart C, eds. *Central Cancer Registries: Design, Management, and Use.* Reading, England: Harwood Academic Publishers; 1994:259–290.
5. Tucker T, Friedell G. Using cancer registry data in primary care practice. *Prim Care Cancer*. 1994;14:33–36.
6. Armstrong B. The role of the cancer registry in cancer control. *Cancer Causes Control*. 1992;3: 569–579.
7. Muir C, Demaret E. The cancer registry in cancer control: an overview. In: Parkin D, Wagner G, Muir C, eds. *The Role of the Registry in Cancer Control.* Lyons, France: International Agency for Research on Cancer; 1985:13–26.
8. Parkin D, Wagner G, Muir C, eds. *The Role of the Registry in Cancer Control*. Lyon, France: International Agency for Research on Cancer; 1985.
9. Ries L, Hankey B, Harras A, Devesa S. Cancer incidence, mortality and patient survival in the United States. In: Schottenfeld D, Fraumeni J, eds. *Cancer Epidemiology and Prevention*. New York, NY: Oxford University Press Inc; 1996:168–191.
10. *SEER Cancer Statistics Review, 1973–1992: Tables and Graphs*. Bethesda, Md: National Cancer Institute; 1996.
11. *Cancer Incidence in North America, 1988–1992.* Sacramento, Calif: North American Association of Central Cancer Registries; 1996.
12. Muir C, Nectoux J. International patterns of cancer. In: Schottenfeld D, Fraumeni J, eds. *Cancer Epidemiology and Prevention.* New York, NY: Oxford University Press Inc; 1996: 141–167.
13. Baquet C, Clayton L, Robinson R. Cancer prevention and control. In: Jones L, ed. *Minorities and Cancer*. New York, NY: Springer-Verlag; 1989:67–76.
14. Boyd N. Epidemiology of cancer. In: Tannock IF, Hill RP, eds. *The Basic Science of Oncology*. New York, NY: McGraw-Hill International Book Co; 1992:7–22.
15. Li F, Kantor A. Cancer epidemiology. In: Holland J, Bast R, Morton D, Frei E, Kufe D, Weichselbaum R, eds. *Cancer Medicine*. Baltimore, Md: Williams & Wilkins; 1997:401–420.
16. Miller BA, Kolonel LN, Bernstein L, et al. *Racial/Ethnic Patterns of Cancer in the United States 1988–1992*. Bethesda, Md: National Cancer Institute; 1996.
17. Centers for Disease Control and Prevention. Cancer registries: the foundation for comprehensive cancer control [At-a-Glance, 1999]. Available at: http://www.cdc.gov/nccdphp/dcpc/npcr/register.htm.
18. Centers for Disease Control and Prevention. State cancer registries: status of authorizing legislation and enabling regulations—United States, October 1993. *MMWR Morb Mortal Wkly Rep.* 1994;43:71–75.
19. Seiffert J. *Standards for Cancer Registries*. Sacramento, Calif: North American Association of Central Cancer Registries; 1997.
20. Seiffert J, Young J. The North American Association of Central Cancer Registries. In: Hutchison C, Roffers S, Fritz A, eds. *Cancer Registry Management: Principles and Practice*. Dubuque, Iowa: Kendall/Hunt Publishing Co; 1997: 375–382.
21. Centers for Disease Control and Prevention. Use of race and ethnicity in public health surveillance: summary of the CDC/ATSDR workshop. *MMWR Morb Mortal Wkly Rep.* 1993; 42(RR-10):1–17.
22. Evinger S. How shall we measure our nation's diversity? *Chance*. 1995;8:7–14.
23. Hahn RA, Stroup DF. Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. *Public Health Rep.* 1994;109:7–15.
24. McKenney N, Bennett CE. Issues regarding data on race and ethnicity: the Census Bureau experience. *Public Health Rep.* 1994;109: 16–25.
25. Osborne NG, Feit MD. The use of race in medical research. *JAMA*. 1992;267:275–279.
26. Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. *BMJ.* 1994;309: 327–330.
27. Kwong S, Perkins C, Snipes K, Wright W. Improving American Indian cancer data in the California Cancer Registry by linkage with the Indian Health Service. *J Registry Manage.* 1998; 23:17–20.
28. Briggs V. *Mass Immigration and the National Interest*. Armonk, NY: ME Sharp Inc; 1996.
29. Johnson-Webb K, Johnson J. North Carolina communities in transition: an overview of Hispanic in-migration. *North Carolina Geographer*. 1996;5:21–40.
30. Lopez L. *Changes in the NC Hispanic/Latino Population, 1990 to 1994—With Implications for Public Health*. Raleigh, NC: Office of Minority Health, Dept of Environment, Health, and Natural Resources; 1995.
31. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston, Mass: Little Brown & Co Inc; 1987.
32. Mills P, Yang R. Cancer incidence in the Hmong of central California, United States, 1987–94. *Cancer Causes Control*. 1997;8:705–712.
33. Aldrich T, Vann D, Moorman P, Newman B. Rapid reporting of cancer incidence in a population-based study of breast cancer: one constructive use of a central cancer registry. *Breast Cancer Res Treat.* 1995;35:61–64.
34. Newman B, Moorman P, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat.* 1995;35: 51–60.
35. Millikan R, Dressler L, Geradts J, Graham M. The need for epidemiologic studies of in-situ carcinoma of the breast. *Breast Cancer Res Treat.* 1995;35:65–77.
36. Friedell GH, Tucker TC, McManmon E, Moser M, Hernandez C, Nadel M. Incidence of dysplasia and carcinoma of the uterine cervix in the Appalachian population. *J Natl Cancer Inst.* 1992;84:1030–1032.
37. Perkins CI, Morris CR, Wright WE, Young JL. *Cancer Incidence and Mortality in California by Detailed Race/Ethnicity, 1988–1992.* Sacramento, Calif: California Dept of Health Services, Cancer Surveillance Section; 1995.
38. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care*. 1993;31: 732–748.
39. Howe H. *Recommendations for Public Use Files of National Cancer Data*. Sacramento, Calif: North American Association of Central Cancer Registries; 1997.
40. Swan J, Wingo P, Clive R, et al. Cancer surveillance in the US—can we have a national system? *Cancer*. 1998;83:1282–1291.